**COVER PAGE**

**Data Science Toolbox: Python Programming**

**PROJECT REPORT**

(Project Semester January-April 2025)

***“Air Quality Insights: A Data-Driven Analysis of Pollution in Indian Cities”***

Submitted by

Adarsh Tiwari

Registration No …12326500……...

Programme and Section …K23EG………

Course Code ...INT217...........

Under the Guidance of

Jaffar Amim Chacket(UID: 30453)

**(Name of faculty coordinator with U.Id and designation)**

**Discipline of CSE/IT**

**Lovely School of \_\_\_Computing\_**

**Lovely Professional University, Phagwara**

**CERTIFICATE**

This is to certify that ...Adarsh Tiwari........ bearing Registration no. ..12326500....... has completed ..INT217......... project titled, ***“"Sales Insights Dashboard: Multi-Factor Analysis of Retail Performance (2020-2024)"***

**”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

**Signature and Name of the Supervisor**

**Designation of the Supervisor**

**School of Computing**

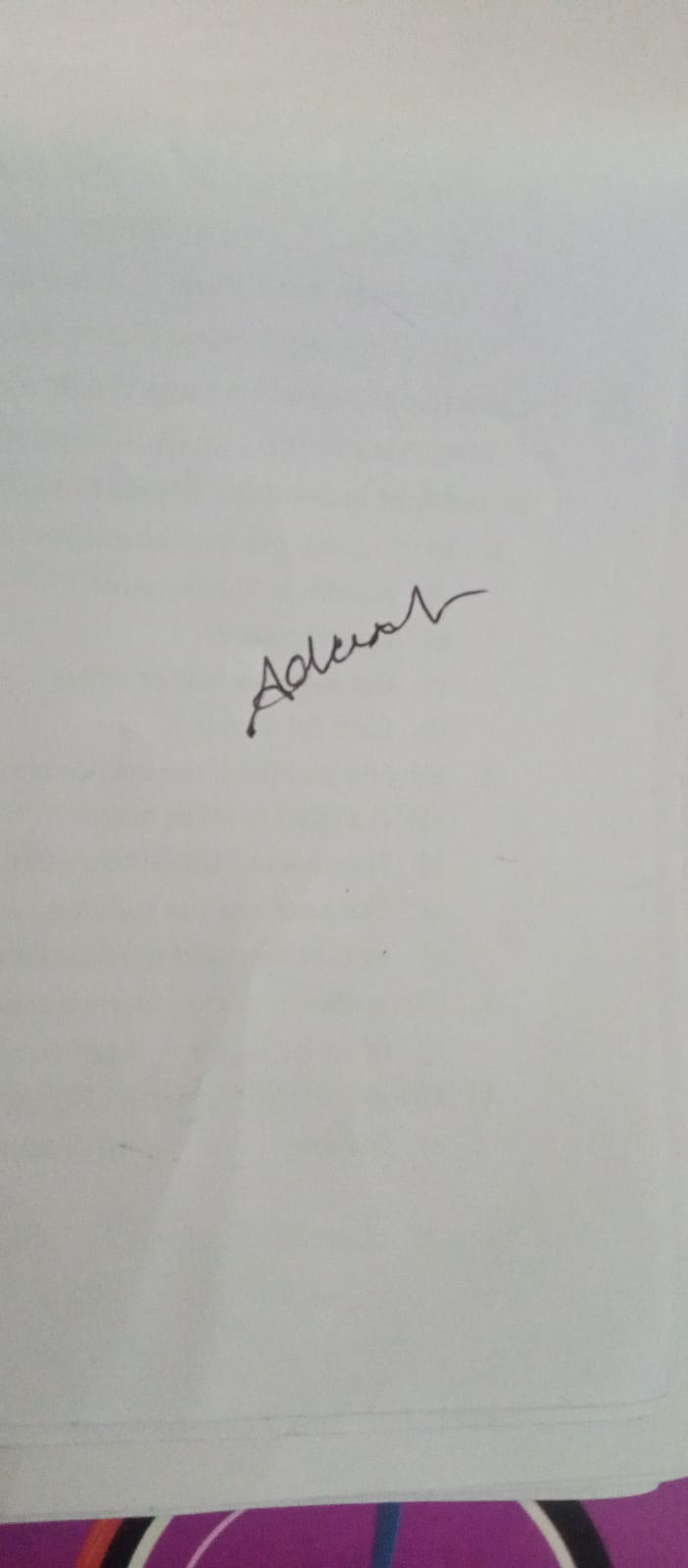
Lovely Professional University

Phagwara, Punjab.

Date: 12/04/2025

**DECLARATION**

I, Adarsh Tiwari, student of BTech CSE (Program name) under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 12/04/2025 Signature: 

Registration No. ..12326500. Name of the student: Adarsh Tiwari

**ACKNOWLEDGEMENT**

I would like to express my heartfelt gratitude to all those who helped me complete this project successfully.

First and foremost, I would like to thank [Faculty Guide's Name], under whose guidance this project was carried out. Their constant support, encouragement, and constructive feedback helped me stay focused and improved the quality of my work.

I would also like to thank the Department of CSE/IT, Lovely Professional University, for providing me with this opportunity and the necessary resources to complete this minor project.

Lastly, I would like to express my appreciation to my family and friends for their unwavering support, motivation, and assistance throughout the duration of this project**.**

**TABLE OF CONTENTS**

| **S. No.** | **Section** | **Page No.** |
| --- | --- | --- |
| **1** | **Certificate** | **\_\_\_** |
| **2** | **Declaration** | **\_\_\_** |
| **3** | **Acknowledgement** | **\_\_\_** |
| **4** | **Introduction** | **\_\_\_** |
| **5** | **Source of Dataset** | **\_\_\_** |
| **6** | **EDA Process** | **\_\_\_** |
| **7** | **Analysis on Dataset** | **\_\_\_** |
|  | **└── Figure 1:Dataset** | **\_\_\_** |
|  | **└── Figure 2: Pivot tables** | **\_\_\_** |
|  | **└── Figure 3: Charts** | **\_\_\_** |
|  | **└── Figure 4 Dashboard** | **\_\_\_** |
|  |
| **8** | **Conclusion** | **\_\_\_** |
| **9** | **Future Scope** | **\_\_\_** |
| **10** | **References** | **\_\_\_** |

**1. Introduction**

**Air pollution has become one of the most pressing environmental challenges in India, with far-reaching consequences for human health, ecological balance, and climate stability. The increasing concentration of harmful pollutants in the air—especially in urban and industrial regions—poses severe threats to both living organisms and the environment. Major metropolitan areas like Delhi, Mumbai, and Kolkata frequently record air quality indices in the "unhealthy" or "hazardous" range, affecting the quality of life for millions.**

**This project aims to conduct a detailed and data-driven analysis of air quality across various Indian cities and states using a dataset compiled from government-approved air quality monitoring stations. The dataset, recorded as of 26th March 2024 at 22:00, contains information on geographic location, pollutant types, and numerical values for minimum, maximum, and average concentrations.**

**The pollutants analyzed in this study include:**

* **PM2.5: Fine particulate matter with a diameter of 2.5 micrometers or less, capable of penetrating deep into the lungs.**
* **PM10: Inhalable particles less than 10 micrometers in diameter, known to cause respiratory issues.**
* **SO₂ (Sulfur Dioxide): A gas resulting from the burning of fossil fuels, contributing to acid rain and respiratory problems.**
* **CO (Carbon Monoxide): A colorless, odorless gas that interferes with the oxygen-carrying capacity of the blood.**

**Through this analysis, the study aims to:**

* **Identify the most and least polluted cities and states in India.**
* **Determine statistical patterns in pollutant concentrations.**
* **Understand how pollution is distributed geographically.**
* **Visualize pollutant behavior using bar plots, pie charts, scatter plots, boxplots, heatmaps, and more.**
* **Provide insight into areas that require immediate governmental or community-level interventions.**

**By leveraging modern data visualization and analytical techniques, this project not only uncovers key environmental insights but also contributes to the broader dialogue on sustainability and public health in India. The findings can aid policy-makers, environmental agencies, and researchers in designing targeted interventions and tracking the effectiveness of pollution control initiatives over time.**

**2. Source of Dataset**

**The dataset used in this project was sourced from Kaggle, a popular open-data community platform that hosts real-world datasets across industries. The data pertains to a retail company's transactional records from 2020 to 2024, and includes fields such as order date, revenue, sales volume, product category, order status, and customer and regional information. This dataset was ideal for performing sales, product, and customer-centric analysis using Excel-based dashboards.**

**3. EDA Process (Exploratory Data Analysis)**

**Exploratory Data Analysis (EDA) is crucial for understanding data structure, identifying trends, and detecting anomalies. In this project, EDA was conducted using Excel’s built-in functionalities, focusing on summarization, filtering, and aggregation to uncover key patterns in the sales dataset.**

**3.1 Data Loading and Initial Exploration**

* **Reviewed structure of the dataset by inspecting column types and sample entries.**
* **Identified key attributes: order date, category, customer ID, region, order status, quantity, and revenue.**
* **Checked for blank or inconsistent entries.**

**3.2 Data Cleaning and Preprocessing**

* **Removed duplicate rows and blank entries.**
* **Standardized date formats and ensured numerical fields (quantity, revenue) were correctly typed.**
* **Created helper columns such as “Year” from the order date for trend analysis.**

**3.3 Statistical Summary**

* **Calculated descriptive statistics (sum, average, min, max) for revenue and quantity.**
* **Grouped data by year, region, and category for comparative summaries.**

**3.4 Group-wise Analysis**

* **Revenue and sales grouped by product category, region, and customer ID.**
* **Customer segmentation based on average order value and purchase frequency.**
* **Analyzed trends in order status distribution (completed, pending, cancelled).**

**3.5 Geographical Analysis**

* **Sales aggregated by region and country.**
* **Created pivot tables to rank regions based on total revenue and total orders.**

**3.6 Data Visualization To aid interpretation, multiple Excel-based visualizations were created:**

* **Line Chart: Year-wise revenue trend (2020–2024)**
* **Column Chart: Product category performance (revenue & quantity)**
* **Pie Chart: Order status breakdown**
* **Bar Graph: Top-performing products by quantity and revenue**
* **Stacked Chart: Regional revenue contributions**
* **Pivot Charts: Customer segments by AOV and frequency**

**3.7 Insight Generation Each visualization led to key takeaways:**

* **Detected seasonal and annual growth trends.**
* **Identified top-selling product categories and best-performing regions.**
* **Highlighted customer segments with highest contribution to sales.**
* **Analyzed operational efficiency through order status proportions.**

**4. Project Objectives & Analytical Approach**

**The goal of this project was to derive business insights from retail sales data using Excel’s powerful capabilities. The core objectives were:**

1. **Analyze Revenue Trends (2020-2024):**
   * **Tracked yearly revenue trends to identify growth or decline.**
   * **Used line charts to visualize and compare year-over-year changes.**
2. **Evaluate Product Category Performance:**
   * **Compared revenue and sales volume across product categories.**
   * **Used bar charts and stacked column visuals to show comparative performance.**
3. **Segment Customers by Purchase Behavior:**
   * **Analyzed purchase frequency and average order value per customer.**
   * **Created customer segments for identifying high-value and frequent buyers.**
4. **Visualize Order Status Breakdown:**
   * **Visualized the distribution of order statuses (Completed, Pending, Cancelled).**
   * **Enabled operational insights and supply chain efficiency tracking.**
5. **Geographical Sales Distribution:**
   * **Mapped regional and country-level sales to find high-performing and underperforming markets.**
   * **Used pivot charts for visual clarity.**
6. **Identify Top-Performing Products:**
   * **Ranked products by revenue and quantity sold.**
   * **Identified potential bestsellers and low-performing SKUs.**

**5. Tools Used**

* **Microsoft Excel (Pivot Tables, Pivot Charts, Slicers)**
* **Data Cleaning in Excel (Removing blanks, transforming date fields)**
* **Kaggle Dataset (Retail sales, product, and customer data)**

|  |  |
| --- | --- |
| 1. **Visualization:** |  |
|  |  |
|  |  |

**5. Conclusion**

This Excel-based dashboard project enabled efficient and visually appealing analysis of retail sales data across multiple dimensions. Revenue trends revealed significant year-on-year growth in specific product categories and regions. Customer segmentation helped identify the most valuable customer groups based on frequency and spending. Geographical and status-based breakdowns highlighted fulfillment challenges and regional opportunities.

Overall, this project demonstrates how traditional tools like Excel can still deliver deep business intelligence when used effectively. It also builds a strong foundation for future exploration in business analytics, especially using advanced BI tools.

**6. Future Scope**

* Integrate dynamic dashboards with real-time data.
* Apply forecasting methods for future revenue predictions.
* Include promotional and discount impact on product performance.
* Expand to Power BI or Tableau for scalable reporting.
* Merge customer feedback or satisfaction data for holistic analysis.

**7. References**

* **Dataset Source: Kaggle (Retail Sales Data 2020–2024)**
* **Microsoft Excel Documentation**
* **Business Analytics Resources (online guides and tutorials)**
* **Visualization Best Practices (Storytelling with Data)**